COMP9313 2018s2 Assignment

z5092923
Jintao Wang

Question 1

1.Firstly, build a neighbour dictionary (say it **Neighbours**) to store the neighbours of each vertex.
2.Then emit all the edges according to the **Neighbours** (map) and remove duplicates (reduce), say the result of this step is **Edges**.
3.Then, for each edge in **Edges**, enumerate all the possible edges connecting to this edge according to **Neighbours**.
4.Finally, rank the results in ascending order by comparing the first node and then the second node and remove edges where first node is larger than the second node.

Question 2

(i)2-shingles for A: S(A) = {the sky; sky is; is blue; blue the; the sun; sun is; is bright}
2-shingles for B: S(B) = {the sun; sun in; in the; the sky; sky is; is bright}
$Similarity = (S(A) \cap S(B))/S(A) \cup S(B)$ = 4 / 9 = 0.45
(ii) M = 22
Input matrix:

	Token	A	B	h1	h2
0	the sky	1	1	8	1
1	sky is	1	1	4	3
2	is blue	1	0	0	5
3	blue the	1	1	5	7
4	the sun	1	1	1	0
5	sun is	1	1	6	2
6	is bright	1	1	2	4
7	sun in	1	0	7	6
8	in the	1	1	3	8

Signature matrix:

A	B
0	1
0	0